Heuristics for Assessment of PID Control with Ziegler-Nichols Tuning

Åström, Karl Johan; Hang, Chang C.; Persson, Per

1988

CODEN: LUTFD2/(TFRT–7404)/1–20/(1988)

Heuristics for Assessement of PID control with Ziegler-Nichols Tuning

K. J. Åström, C. C. Hang, and P. Persson

Department of Automatic Control
Lund Institute of Technology
November 1988

Department of Automatic Control
Lund Institute of Technology
P.O. Box 118
S-221 00 Lund Sweden

Document name
Technical report

Date of issue
November 1988

Document Number
CODEN: LUTFD2/(TFRT–7404)/1–20/(1988)

Author(s)
K. J. Åström
C. C. Hang
P. Persson

Supervisor

Sponsoring organisation
STU, The Swedish Board for Technical Development under contract DUP 85-3084P.

Title and subtitle
Heuristics for Assessement of PID control with Ziegler-Nichols Tuning

Abstract

In this paper we attempt to develop formal tools to assess what can be achieved by PID control of a class of systems with the Ziegler-Nichols tuning formula and to characterize a class of systems where PID control is appropriate. Based on empirical results and approximate analytical study, we introduce two numbers, namely the normalised dead time θ and the normalized process gain κ, to characterize the open loop process dynamics and two numbers, the peak load error λ and the normalised rise time τ, to characterize the closed loop response. Simple methods of measuring these parameters are proposed.

It is shown that θ and κ are related and either of them can be used to predict the achievable performance of PID controller tuned by the Ziegler-Nichols formula. A small θ indicates that tight control is achievable with P or PI control. Processes with θ in the range of 0.15 to 0.6 can be controlled well with PID regulators. Moderate performance can only be expected if θ is larger than 0.6 and hence a more sophisticated controller like Smith Predictor should be used for tight control. The intelligent controller can thus interact with the operator and advise on choice of control algorithm.

We have established useful relations, such as $\tau \approx 1$ and $\kappa\lambda \approx 1.3$, which can be used to assess whether the PID controller is properly tuned. The simplicity of the relations allows the development of a first generation of intelligent controller using current technology.

Key words

Classification system and/or index terms (if any)

Supplementary bibliographical information

ISSN and key title

ISBN

Language
English

Number of pages
20

Recipient's notes

Security classification

Heuristics for Assessment of PID control with Ziegler-Nichols Tuning

K. J. Åström, C. C. Hang, and P. Persson

Abstract.

In this paper we attempt to develop formal tools to assess what can be achieved by PID control of a class of systems with the Ziegler-Nichols tuning formula and to characterize a class of systems where PID control is appropriate. Based on empirical results and approximate analytical study, we introduce two numbers, namely the normalised dead time θ and the normalized process gain κ, to characterize the open loop process dynamics and two numbers, the peak load error λ and the normalised rise time τ, to characterize the closed loop response. Simple methods of measuring these parameters are proposed.

It is shown that θ and κ are related and either of them can be used to predict the achievable performance of PID controller tuned by the Ziegler-Nichols formula. A small θ indicates that tight control is achievable with P or PI control. Processes with θ in the range of 0.15 to 0.6 can be controlled well with PID regulators. Moderate performance can only be expected if θ is larger than 0.6 and hence a more sophisticated controller like Smith Predictor should be used for tight control. The intelligent controller can thus interact with the operator and advise on choice of control algorithm.

We have established useful relations, such as $\tau \approx 1$ and $\kappa\lambda \approx 1.3$, which can be used to assess whether the PID controller is properly tuned. The simplicity of the relations allows the development of a first generation of intelligent controller using current technology.

1. Introduction

The thrust of control theory for the past 30 years has been to provide exact solutions to precisely stated problems. Much less work has been devoted to finding crude solutions to poorly defined problems. One of the few exceptions is the work on fuzzy sets by Zadeh (1973). A typical example is control system design where a lot of prior knowledge like a mathematical model, design criteria etc. is required to carry out a design. To make a good control system design it is also very useful to have an assessment of some key features of the system like bandwidth, achievable performance, etc. There are also several problems, like integral windup etc, that have to be handled. This is normally done manually by engineers. Not much is published about these craft-like aspects of control. It is certainly not part of the standard control curriculum. This has undoubtedly contributed to the recurrent discussions on the gap between theory and practice in control.

Why should we then be concerned with these issues? We have been led into this in efforts to design expert control system (Åström et. al. 1986) where some of the knowledge of design engineers is built into a control system. We seek to extract and condense knowledge about control system design which can replace the otherwise large number of possibly conflicting rules accumulated by different experts to ease the workload of a real-time expert system. We also believe that it is useful to describe the heuristic aspects of control so that the knowledge can be discussed and refined. This can also contribute to spreading control engineering knowledge to persons with less formal education. A long range goal is to provide a framework for making qualitative reasoning about control systems.

This paper looks at a simple version of the problem. It tries to give formal tools to assess what can be achieved by PID control of a restricted class of processes and a simple tuning rule. The key result is that there are simple dimension-free parameters that give insight into the achievable performance. These features will allow us to do formal reasoning about simple control loops.

The paper is organized as follows. The restricted class of processes that we are concerned with is introduced in Section 2. Some useful dimensionless numbers are introduced in Section 3. In Sections 4 and 5 some relations between the features are derived by approximate analysis and empirical refinement based on simulation. The results are used in Section 6 to discuss the performance that may be achieved with PID control based on Ziegler-Nichols tuning. A possible application of the process characteristics in detecting instrumentation errors is outlined in Section 7. Some conclusions are given in Section 8.

2. Process Characteristics

The processes we consider are restricted to simple feedback loops. It is assumed that the process dynamics is linear and stable. The characteristics will be further restricted both in the time and the frequency domain.

2.1 Time Domain Characterization

It will be assumed that the step response has the general characteristics shown in Figure 2.1.

A system with a positive impulse response clearly has a monotone step response. The fact that the impulse response is unimodal ensures that the step response has a unique inflexion point. An impulse response is essentially positive if it is positive possibly apart from a small initial part. This is the essential feature that we will use because for such systems the quantities k_p, L, and T can be defined. The number k_p is the *static process gain*, the number L is the *apparent dead time* and the number T is the *apparent time constant*. The parameters T and L are obtained by the graphical construction indicated in Figure 1 where the tangent is drawn in the inflexion point of the step response. An alternative is to define L by drawing a line between the points where the step response has reached 10% and 90% of its steady state values. The transfer function

$$G(s) = k_p \frac{e^{-sL}}{1+sT} \tag{2.1}$$

is a crude analytic approximation of the the transfer function of the class of processes that we are considering. Notice however that the transfer functions considered are not restricted to this class. It is sufficient to have a step response with the shapes shown in Figure 2.1a or Figure 2.1b, characterized by k_p, L, and T.

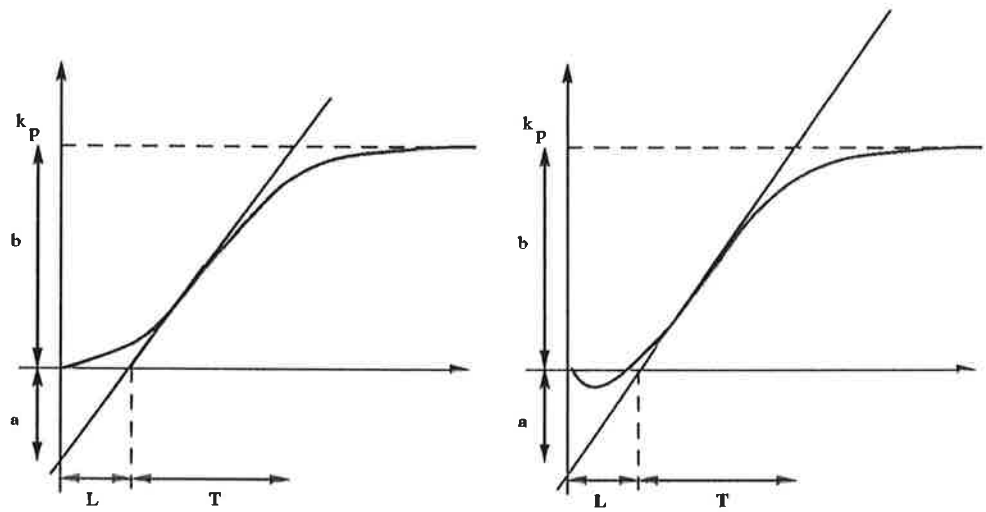

Figure 2.1 Step response of a system whose impulse response is a) positive and unimodal and b) essentially positive and unimodal.

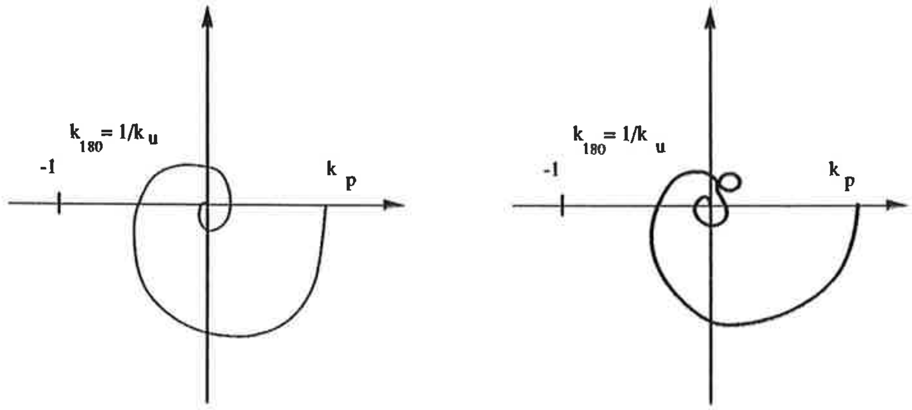

Figure 2.2 Nyquist curve for system with a) monotone frequency response and b) essentially monotone frequency response.

The class of systems considered is the same as that used in the classical works on Ziegler-Nichols tuning. There are important classes of systems that are excluded, e.g., systems having integrators and systems with resonant poles. Systems having integrators may have monotone step responses but they are not stable. Systems with resonant poles do not have a monotone step response.

2.2 Frequency domain characterization

A different frequency domain characterization of process dynamics will also be introduced. It is assumed that the Nyquist curve has the shape indicated in Figure 2.2.

To be specific it is assumed that both the phase and the amplitude are monotone functions of the frequency. This guarantees that the the intersections with the real and imaginary axes are unique. The first intersection with the negative real axis defines the *ultimate frequency*, ω_u, and the *ultimate gain*, k_u. Lack of monotonicity can be accepted at high frequencies.

3. Features

Dimension-free parameters, like Reynold's numbers, have found much use in many branches of engineering. They have however not been much used in automatic control. In this section it is attempted to introduce some numbers that are useful in assessing control system performance.

3.1 Normalized Dead-time

The normalized dead-time is defined as the ratio of the apparent dead-time and the apparent time constant, or formally

$$\theta = \frac{L}{T} = \frac{a}{k_p}. \tag{3.1}$$

See Figure 2.1. This number is thus easily obtained from a record of the step response. It has been known from practical experience that the normalized dead-time may be used as a measure of the difficulty of controlling a process. Processes with a small θ are easy to control and processes with a large θ are difficult to control. The parameter θ was actually called the *controllability ratio* by Deshpande and Ash (1981). Fertik (1975) introduced the name *process controllability* for the quantity $\theta/(1+\theta)$. To avoid possible confusion with the standard terminology of modern control theory we will use the word *normalized dead time.*

3.2 Normalized Process Gain

The process gain k_p is not dimension-free. It can however be made dimension free by multiplication with a suitable regulator gain. The ultimate gain k_u, i.e., the regulator gain that makes the process unstable under proportional feedback control, is a suitable normalization factor. With reference to Figure 2.2 the normalized process gain, κ, can thus be defined as

$$\kappa = k_p k_u. \tag{3.2}$$

This number is easily obtained as the ratio of the process gains where the phase is 0° and 180°, see Figure 2.2. The number also has a physical interpretation as the largest process loop gain that can be achieved under proportional control. The number is useful to assess the control performance. Roughly speaking, a large value indicates that the process is easy to control while a small value indicates that the process is difficult to control.

The normalized process gain is directly obtained from a Nyquist curve of the process. It can also be obtained from an experiment with relay feedback, see Åström and Hägglund (1984).

Since the processes we consider are stable they have a static error under proportional feedback. The static error obtained for a unit step command

$$e_s = \frac{1}{1+k_pk_c} > \frac{1}{1+\kappa} \tag{3.3}$$

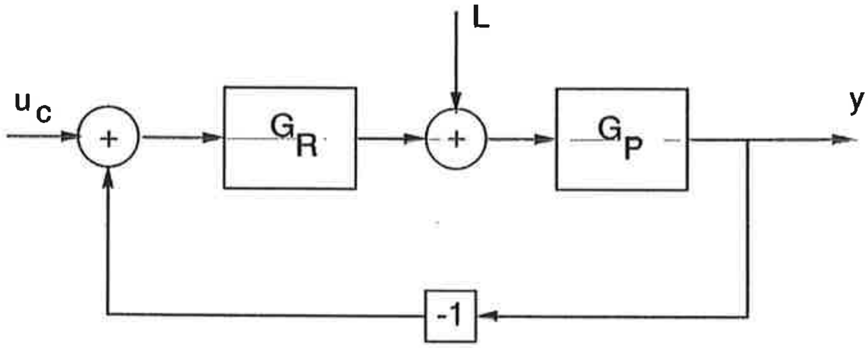

Figure 3.1 Block diagram of a simple feedback system with a load disturbance acting at the process input.

where k_c is the proportional gain used. The inequality follows because $k_p k_c < \kappa$. The number κ can thus be used to estimate the static error achievable under proportional control and also to determine if integral action is required to satisfy the specifications on static error.

3.3 Peak Load Disturbance Error

The response to step load disturbances is an important factor when evaluating control systems. The effect of a load disturbance depends on where the disturbance acts on the system. In this section it will be assumed that the disturbance acts on the process input, see Figure 3.1.

With a regulator without integral action a unit step disturbance in the load gives the static error

$$e_l = \frac{k_p}{1 + k_u k_c} > \frac{k_p}{1 + \kappa}. \tag{3.4}$$

The quantity e_l / k_p is dimension-free.

When a regulator with integral action is used the static error due to a step load disturbance is zero. A meaningful measure is then the maximum error due to a load disturbance. To obtain a dimension-free quantity it is also divided by the process gain. The following variable is thus obtained

$$\lambda = \frac{1}{l_0 k_p} \max e(t) \tag{3.5}$$

where l_0 is the amplitude of the step disturbance.

3.4 Normalized Closed Loop Rise Time

The closed loop rise time is a measure of the response speed of the closed loop system. Again, to obtain a dimension-free parameter it will be normalized by the apparent dead time L of the open loop system. The parameter is thus

$$\tau = \frac{t_r}{L}. \tag{3.6}$$

4. Empirics

The Ziegler-Nichols closed-loop tuning procedure was applied to a large number of different processes. It was attempted to correlate the observed properties of the open and closed loop systems to the features introduced in Section 3. In this section we will present the empirical results. Processes with the transfer functions

$$G_1(s) = \frac{e^{-sD}}{(1+s)^2} \tag{4.1}$$

$$G_2(s) = \frac{1}{(1+s)^n}, \quad 3 < n < 20 \tag{4.2}$$

$$G_3(s) = \frac{1-\alpha s}{(1+s)^3}, \quad 0 < \alpha < 2.5 \tag{4.3}$$

will be investigated. These models cover a wide range of dynamic characteristics such as pure dead-time and nonminimum phase response. The main features of the models are summarized in Appendix A.

The normalized apparent dead-time was measured from the step responses. The ultimate gain was determined by simulation. Parameters of PID regulators were determined by a straight forward application of the Ziegler-Nichols closed-loop method without fine tuning, i.e. with values of proportional gain k_c, integral time T_i and derivative time T_d set as $0.6k_u$, $0.5T_u$, and $0.125T_d$ respectively. The closed loop performance is judged based on the closed loop step and load responses.

The results obtained are summarized in Tables 1–3. The tables give a parameter that characterizes the process, the ultimate period T_u, the overshoot os, the undershoot us of the closed loop step response, the apparent normalized dead-time $\theta = L/T$, the normalized loop gain κ, the product $\kappa\theta$, the normalized closed loop rise time $\tau = t_r/L$, the normalized peak load error λ, and the product $\omega_u t_r$.

The results for the first process are summarized in Table 1. The closed loop behaviour was judged to be satisfactory for $0.15 < \theta < 0.6$. The overshoot for θ in the low range is too high. This is however easily reduced by using the setpoint weighting factor modification, see Åström and Hägglund (1988). For large values of θ there is a pronounced undershoot in the step response.

D	T_u	os	us	θ	κ	$\kappa\theta$	τ	λ	$\kappa\lambda$	$\omega_u t_r$
0.1	1.4	75	26	0.15	21	3.2	0.80	0.06	1.26	1.5
0.2	2.0	60	14	0.19	10.5	2.0	0.95	0.15	1.57	1.5
0.4	2.8	50	5	0.26	5.7	1.5	1.0	0.27	1.53	1.6
0.6	3.6	35	2	0.34	4.0	1.4	0.94	0.37	1.48	1.7
1.0	4.8	26	3	0.49	2.7	1.3	1.02	0.52	1.40	1.7
1.5	6.0	19	9	0.69	2.0	1.4	0.93	0.66	1.35	1.8
2.0	7.2	14	14	0.89	1.7	1.5	0.85	0.73	1.26	1.8
2.5	8.3	12	17	1.09	1.5	1.6	0.82	0.83	1.28	1.8
3.0	9.4	20	20	1.26	1.4	1.8	0.79	0.89	1.25	1.8

Figure 4.1 Table 1. Experimental results for a system with the transfer function $G(s) = e^{-sD}/(s+1)^2$.

The results for the second process are summarized in Table 2. The closed loop behaviour was judged to be satisfactory for $0.22 < \theta < 0.64$. The overshoot for θ in the low range is too high. This is however easily reduced by using the setpoint weighting factor modification. For large values of θ there is a pronounced undershoot in the step response. Similar results are obtained for the third process as summarized in Table 3.

n	T_u	os	us	θ	κ	$\kappa\theta$	τ	λ	$\kappa\lambda$	$\omega_u t_r$
3	3.7	50	13	0.22	8.0	1.52	1.07	0.19	1.52	1.3
4	6.0	40	10	0.32	4.0	1.4	1.16	0.35	1.40	1.7
6	10.6	26	11	0.49	2.4	1.29	1.14	0.54	1.30	1.9
8	14.6	17	14	0.64	1.88	1.18	1.08	0.64	1.18	2.0
10	18.8	12	17	0.76	1.60	1.18	0.96	0.74	1.18	
15	29.0	0	24	1.05	1.36	1.15	0.9	0.85	1.16	?
20	39.0	5	30	1.28	1.25	1.14	0.8	0.91	1.14	?

Figure 4.2 Table 2. Experimental results for a system with the transfer function $G(s) = 1/(s+1)^n$.

α	T_u	os	us	θ	κ	$\kappa\theta$	τ	λ	$\kappa\lambda$	$\omega_u t_r$
0	3.7	50	13	0.22	8	1.54	1.15	0.19	1.52	1.3
0.1	3.8	50	15	0.23	6.2	1.49	1.09	0.24	1.49	1.5
0.25	4.3	48	11	0.28	4.5	1.44	1.09	0.32	1.44	1.7
0.5	5.0	38	3.8	0.38	3.2	1.41	1.16	0.44	1.41	2.0
1.0	6.0	21	3.8	0.58	2.0	1.34	0.98	0.67	1.34	2.2
1.5	6.5	9.6	7.7	0.76	1.45	1.31	0.89	0.90	1.30	2.4
2.0	7.0	-1.9	16	0.98	1.15	1.24	0.84	1.08	1.24	2.7

Figure 4.3 Table 3. Experimental results for a system with the transfer function $G(s) = (1-\alpha s)/(s+1)^3$.

5. Relations

We have thus introduced two normalized numbers, namely the normalized dead-time θ and the normalized process gain κ, to characterize the open loop dynamics and two numbers, the peak load error λ and the normalized closed loop rise time τ to characterize the closed loop response. Some relations between these numbers will now be established. In doing so we will also develop an intuitive feel for the meaning of the numbers. The prototype for our reasoning is the well known relation between bandwidth and rise time for an electronic amplifier. A relation will first be derived mathematically using several approximations. This gives the possible mathematical form of the relation. A number of specific examples will then be solved to find the numerical parameters of the coefficients of the relation. Since the equations which we are searching relate open loop and closed loop properties they will depend on the regulator structure and the design method. Throughout the paper it will be assumed that Ziegler-Nichols tuning is used.

5.1 Rise Time Bandwidth Product

In the design of electronic amplifiers it has been noticed that the product of the bandwidth and the rise time is approximately constant. This can be derived as follows. Let $G(s)$ be the closed loop transfer function of the amplifier and $H(t)$ the unit step response. It follows that

$$\frac{dH}{dt} = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} e^{-st}G(s)\,ds. \tag{5.1}$$

If the rise time t_r is defined as in Figure 2.1 using the maximum slope of the step response we get

$$t_r \max_{0\leq t\leq\infty} \mid \frac{dH}{dt} \mid = H(\infty) = G(0). \tag{5.2}$$

Hence

$$t_r \int_0^\infty \mid \frac{G(i\omega)}{G(0)} \mid d\omega = \pi. \tag{5.3}$$

The integral on the left hand side is approximately equal to the bandwidth ω_b of the system. Summarizing we find the following relation between rise time and bandwidth

$$t_r\omega_b \approx \pi. \tag{5.4}$$

With Ziegler-Nichols tuning and PID control the bandwidth of the closed loop system is approximately proportional to the ultimate frequency ω_u. We can thus expect that the product $t_r\omega_u$ is constant. The empirical results obtained in the previous section also supports this and we get

$$t_r\omega_u \approx 2. \tag{5.5}$$

Compare with Tables 1–3.

5.2 Normalized Dead-time and Process Gain

As can be seen from Tables 1–3 there appears to be a relation between normalized process gain κ and normalized dead time θ. For specific systems it is possible to find the relations exactly, see the appendices. For first order systems with dead time we have:

$$\theta = \frac{\omega_u L}{\omega_u T} = \frac{\pi - \arctan\sqrt{\kappa^2 - 1}}{\sqrt{\kappa^2 - 1}} \tag{5.6}$$

See appendix A. This relation is shown graphically in Figure 5.1.

It is possible to find exact expressions for the relations between κ and θ for the processes given by equations (4.1), (4.2) and (4.3). They can also be obtained experimentally as discussed in Section 4. The relations are shown in Figure 5.1. The graphs indicate that for processes with higher order dynamics the product $\kappa\theta$ is approximately constant. This is important because it means that the normalized process gain κ can be used instead of the normalized dead time θ to assess achievable performance.

It is also interesting to note that the curve for the exact model given by equation (5.6) deviates substantially from those of the higher order models, particularly in the region $0.3 < \theta < 1.2$. The normalized gain can be as much as 0.5 to 1 unit higher for the same θ in this range. If both κ and θ are determined this information can be used to assess if the minimum phase part of the dynamics is first order or not.

Apart from application as diagnostics to indicate whether the Ziegler-Nichols tuning will work well the relation between κ and θ can be used for fine tuning. For instance it is known that the Ziegler-Nichols tuning uses a ratio $T_i/T_d = 4$ which gives good response to load disturbance for processes with high order dynamics. See Hang (1989). This ratio should be higher for processes with first order dynamics and lower for processes with oscillatory poles. This observation opens yet another possibility to incorporate intelligence in the controller.

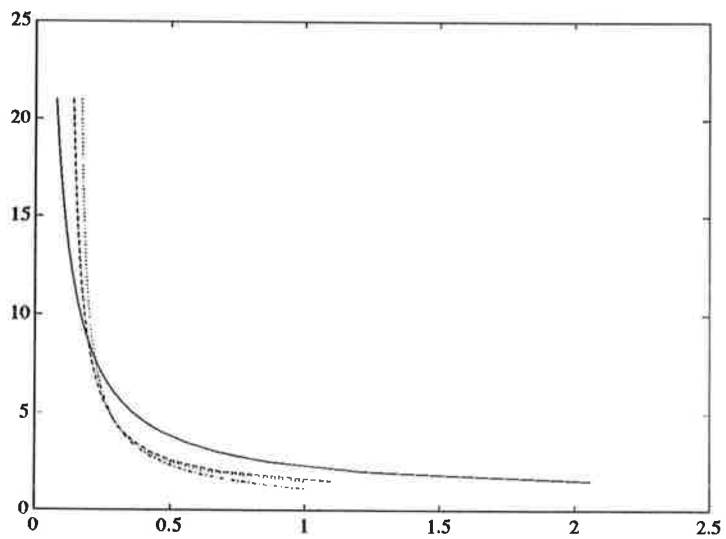

Figure 5.1 The normalized process gain κ as a function of apparent normalized dead time θ for systems (2.1), (4.1), (4.2), and (4.3).

5.3 Peak Load Error and Normalized Dead-time

Consider the closed loop system obtained with the process and the regulator. Assume that the disturbance enters at the plant input. The transfer function from the load disturbance to the output is

$$G_d(s) = \frac{G_p(s)}{1 + G_p(s)G_r(s)} = \frac{1}{G_r(s)} \frac{G_p(s)G_r(s)}{1 + G_p(s)G_r(s)}. \tag{5.7}$$

A PID regulator with Ziegler-Nichols tuning has the transfer function

$$G_r(s) = \frac{k_r(s+a)^2}{2as} \tag{5.8}$$

where

$$a = \frac{1}{2T_d} = \frac{4}{T_u}. \tag{5.9}$$

This choice ensures near optimal load rejection, as discussed by Hang (1989). With Ziegler-Nichols tuning the closed loop system has a time constant $T_r = 0.85T_u$ which corresponds to a bandwidth of $\omega = 7.4/T_u$. From (5.11) we thus get the following approximate formula

$$G_d(s) \approx \frac{1}{G_r(s)} = \frac{2as}{k_r(s+a)^2}. \tag{5.10}$$

The corresponding unit step response is

$$H(t) = \frac{2at}{k_r} e^{-at} \tag{5.11}$$

which has a maximum

$$\frac{2}{ek_r} = \frac{0.74}{k_r} = \frac{1.23}{k_u} \tag{5.12}$$

at

$$t = \frac{1}{a} = 2T_d. \tag{5.13}$$

Summarizing we find that we can expect the parameter $\kappa\lambda$ to be constant. This is also supported by the experimental results given in Tables 1–3 which gives

$$\kappa\lambda \approx 1.3. \tag{5.14}$$

The knowledge of λ can be used by an intelligent controller to check if a PID controller with Ziegler-Nichols tuning can be used to satisfy the given specifications to peak load error. From the analysis we also find that the peak error occurs $T_u/4$ time units after the step disturbance is applied.

5.4 Closed Loop Rise Time

The experimental results given in Tables 1–3 show that the normalized rise time is approximately constant. Hence

$$\tau \approx 1. \tag{5.15}$$

In physical terms this implies that $t_r \approx L$, compare with equation (3.6). This means that the Ziegler-Nichols method gives a closed loop system with a rise time approximately equal to the apparent dead-time of the open loop system.

6. Ziegler-Nichols Tuning

The results obtained will now be used to evaluate PID regulators with Ziegler-Nichols tuning. We can first observe that the Ziegler-Nichols tuning procedure is very simple. It is based on a simple characterization of the process dynamics, either parameters a and L from the step response or the critical point on the Nyquist curve parameterized in k_u and ω_u. We have also obtained two relations $\tau \approx 1$ and $\kappa\lambda \approx 1.3$ which characterizes the closed loop performance. The condition $\tau \approx 1$ implies that Ziegler-Nichols tuning tries to make the closed loop rise time equal to the apparent dead-time.

6.1 When can Ziegler-Nichols Tuning be used?

The results obtained show that Ziegler-Nichols tuning will give good results under certain conditions and that these conditions can be characterized by one parameter, θ, or $\kappa = k_u k_p$.

The results are summarized in Table 4.

θ	Tight Control is Not Required	Tight Control is Required		
		High Measurement Noise	Low Saturation Limit	Low Measurement Noise and High Saturation Limit
Class I < 0.15	P	PI	PI or PID	P or PI
Class II $0.15 \sim 0.6$	PI	PI	PI or PID	PID
Class III $0.6 \sim 1$	I or PI	I + A	PI + A	PI or PID + A + C
Class IV > 1	I	I + B + C	PI + B + C	PI + B + D

Figure 6.1 Table 4. A: Feedforward compensation recommended, B: Feedforward compensation essential, C: Dead-time compensation recommended, D: Dead-time compensation essential.

Four cases are introduced in the table. They are classified as follows:

Case 1 $\theta < 0.15$ **or** $\kappa > 20$ **:** Ziegler-Nichols tuning may not give the best results in this case. The reason is that it is possible to use comparatively high loop gains. There are many possible choices of regulators. A P or PD regulator may be adequate if the requirements on static errors are not too stringent. A proportional regulator could be chosen if a static error around 10% is tolerable. (This estimate is based on the assumption that the regulator gain is half of the ultimate gain). If smaller static errors are required it is necessary to use integral action. Performance can often be increased significantly by using derivative action or even more complicated control laws. Temperature control where the dynamics is dominated by one large time constant is a typical case. We have observed that the derivative time $T_d = T_i/4$ obtained by the Ziegler-Nichols rule is too large in this case. It gives a long tail in the step response; a better value is $T_d = T_i/8$.

Case 2 $0.15 < \theta < 0.6$ **or** $2 < \kappa < 20$ **:** This is the prime application area for PID controllers with Ziegler-Nichols tuning. It works well in this case. Derivative action is often very helpful.

Case 3 $0.6 < \theta < 1$ **or** $1.5 < \kappa < 2$ **:** When θ approaches 1 Ziegler-Nichols tuning becomes less useful. This is easy to understand if we recall that the tuning procedure tries to make closed loop rise time equal to the apparent dead time. It is difficult to achieve tight control with Ziegler-Nichols tuned PID regulators. Other tuning methods and other regulator structures like Smith predictors, pole placement, or feedforward could be considered.

Case 4 $\theta > 1$ or $\kappa < 1.5$: PID control based on Ziegler-Nichols tuning is not recommended when θ is larger than 1. The reason why the regulators work so poorly for $\theta > 0.6$ is partly due to inherent limitations of PID controllers and partly due to the Ziegler-Nichols tuning procedure. Modifications of the Ziegler-Nichols rule were proposed by Cohen-Coon (1953). By choosing other tuning methods it is however possible to tune PID regulators to work satisfactorily even for $\theta = 10$, see Åström (1988).

A parallel effort by Hang and Åström (1988) has gone further than merely using θ to predict the effectiveness of the Ziegler-Nichols tuning formula. The following modification to eliminate manual fine tuning has been recommended. When $\theta < 0.6$ the main drawback of the Ziegler-Nichols formula is excessive overshoot. This can be overcome by setpoint weighting where the weighting factor is a simple function of θ. When $\theta > 0.6$ the integral time computed by the Ziegler-Nichols formula needs to be modified by a factor which again can be expressed as a simple function of θ. These modifications are essential to obtain high quality PID control without manual fine tuning.

Table 4 indicates that a broad classification of Ziegler-Nichols tuned PID controllers can be made based on the normalized dead-time. This observation is useful if we try to build control systems with decision aids where the instrument engineer or the operator is advised also on regulator selection. Table 4 indicates that such recommendations must be based on interaction with the operator because the choices will depend not only on the process characteristics, i.e. θ or κ, but also on performance requirements such as static errors. If tight control is not required then PI control is often adequate and PID control which is more difficult to tune and more sensitive to noise can be avoided. Notice that the choice may be different if regulators with automatic tuning are available, since it is then easier to use regulators with derivative action.

6.2 Implications for Smart Controllers

There are several simple auto-tuners that are based on the Ziegler-Nichols tuning procedure. A drawback with them is that they provide tuning but that they are unable to reason about the achievable performance. The result of this paper indicates that there is a simple modification. By determining one of the parameters θ or κ it is thus a simple matter to provide facilities so that a simple auto-tuner can select the regulator form P, PI, or PID and also give indications if a more sophisticated control law would be useful. For an auto-tuner based on the transient method this can be achieved by determining not only a and L but also k_p and including a logic based on Table 4. For relay based auto-tuners it is necessary to complement the determination of ω_u and k_u with determination of k_p. This can easily be made from measurement of average values of inputs and outputs in steady state operation. It is also possible to modify the relay tuning so that the static gain is also determined. The accuracy of the tuning formula over a wide range of θ-values can be markedly improved by the use of the correlation formula of Hang and Åström (1988) as discussed above.

6.3 On-line Assessment of Control Performance

The results of this paper can also be used to evaluate performance of feedback loops under closed loop operation. Consider, e.g., the relation (5.15) for the normalized rise time. The rise time can be measured when the set point is changed. If the regulator is properly tuned then the closed loop rise time should be equal to the apparent dead time. If the actual rise time is significantly different, say 50% larger, it indicates that the loop is poorly tuned. This type of assessment is particularly useful when the damping is adequate but the Foxboro's Exact, based on pattern recognition, Bristol (1977), cannot make this kind of judgement.

Similarly the relation (5.17) can be used by introducing a perturbation at the regulator output. If the maximum error deviates from that predicted by (5.18) we can suspect that the loop is poorly tuned.

7. Control System Critiqueing

A good control system performance is achievable provided that the control system design is sound, the instrumentation is adequate and undersized the process input will saturate and a fast response cannot be obtained. This is reflected in a small input saturation threshold and a small process gain. On the other hand an oversized control valve will provide the necessary extra power for rapid response and significantly increase the threshold for input saturation. However an excessive oversizing would result in a very small valve motion in steady state regulation and a poor resolution. This is reflected in a very large process gain and a large input saturation limit. Likewise the process output measurement range or calibration can result in too low or too high a static process gain due to over-ranging or under-ranging. In summary, the knowledge of the process gain can indicate control system limitation due to inadequate instrumentation. This knowledge can be improved by on-line monitoring of actuator saturation.

The normal instrumentation practice is to ensure that static operating conditions are satisfied and that appropriate allowance is given for dynamic performance. A static process gain from actuator input to sensor output of 0.5 to 2 is quite common. If the process gain is lower say 0.1 and θ is small, say 0.1 which implies $\kappa = 15$, the regulator gain will be very high $k_r \approx 90$. A set point change as small as 1.1% will then saturate the actuator. If the actuator is resized such that the static process gain becomes 2 then the regulator gain becomes 4.5 and the actuator will not saturate unless the set point change is larger than 22%. In other words, too high a controller gain should be avoided and if required it should be shifted to the process.

In the examples discussed above it has been assumed that the process output is correctly calibrated. The small process gain is then caused by the under-sized actuator. It may however also be due to the measurement being oversized. For instance, if the full range of the output is $10V$ and the full control range only gives $1V$ the static process gain is 0.1. If the measurement is re-ranged so that full output range is used the process gain is 1. It is of course the task of the instrument engineer to make sure that the instrumentation is properly sized but it is nevertheless useful to have diagnostics that indicates that there may be a problem. A reasonable rule is to determine if the gain is in the range 0.5 to 2. There are systems where higher process gains occur, a typical case being a process with a very long time constant, almost like an integrator, where the normalized process gain may be much higher. This occurs, e.g., in some systems for temperature control.

The static process gain can be measured from an open loop step response or from an experiment with relay feedback. It can also be determined from set point changes in closed loop. In view of the importance of the static gain it is advisable to provide tools for its determination even in simple control systems.

It may be argued that instrumentation problems can be identified by the operator or the instrument engineer when they occur. It is, however, useful to have controllers with facilities to indicate potential problems. It seems quite reasonable that future systems will include a critiqueing system which will advise if the sensors and actuators are appropriately chosen.

8. Conclusions

In this paper it has been attempted to analyze simple feedback loops with PID regulators that are tuned using the Ziegler-Nichols closed loop method. It has been shown that there are some quantities that are useful to assess achievable performance and to select suitable regulators. These quantities are the *normalized process gain* (κ), the *normalized dead-time* (θ), the *normalized closed loop rise time* (τ), and the *peak load error* (λ). Simple methods to determine these parameters have also been suggested.

It has been shown that κ or θ are related and that they can be used to assess the control problem. A small θ indicates that tight control is possible with P or PI control but also that significant improvements may be possible with more sophisticated control laws. Processes with θ in the range from 0.15 to 0.6 can be controlled with PID regulators with Ziegler-Nichols tuning. The results show clearly that Ziegler-Nichols tuning gives poor results when the normalized dead-time θ is larger than 0.6. There are also relations like $\tau \approx 1$ and $\kappa\lambda \approx 1.3$, that may be used to assess the closed loop response time and the load rejection properties. The results indicate that it would be useful to determine at least one of the parameters κ or θ in connection with regulator tuning because these parameters are so important for assessment of achievable performance. Some empirical rules for controller selection and assessment have also been given. Knowledge of θ also allows us to incorporate the modified Ziegler-Nichols formula recommended by Hang and Åström (1988). This can be used at both small and large θ so as to eliminate manual fine tuning for good control performance.

8.1 Acknowledgements

This work has been supported by the Swedish Board for Technical Development (STU) under contract DUP 85-3084P.

9. References

BRISTOL E. H. (1977): "Pattern Recognition: An Alternative to Parameter Identification in Adaptive Control," *Automatica*, **13**, 197–202.

COHEN G. H. and G. A. COON (1953): "Theoretical Consideration of Retarded Control," *Transactions of the ASME*, **75**, 827–834.

DESHPANDE P. B. and R. H. ASH (1981): *Computer Process Control*, ISA, Research Triangle Park, NC, USA.

FERTIK H. A. (1975): "Tuning Controllers for Noisy Processes," *ISA Transactions*, **14**.

GOFF K. W. (1966): "Dynamics in Direct Digital Control, Part I," *ISA Journal*, **13**.

GOFF K. W. (1966): "Dynamics in Direct Digital Control, Part II," *ISA Journal*, **13**, 44–54.

HANG C. C. (1989): "Controller zeros," *IEEE Control systems Magazine, to appear.*

HANG C. C. and K. J. ÅSTRÖM (1988): "Refinements of the Ziegler-Nichols Tuning Formula for PID Auto-tuners," *Proc. ISA Annual Conf. Houston, USA.*

ZADEH L. A. (1973): "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes," *IEEE Transactions of Systems, Man, and Cybernetics.*, **SMC-3**, 28–44.

ZIEGLER J. G. and N. B. NICHOLS (1942): "Optimum Settings for Automatic Controllers," *Transactions of the ASME*, **64**, 759–768.

ÅSTRÖM K. J. (1988): "Dominant Pole Placement Design of PI Regulators," CODEN: LUTFD2/TFRT-7381, Department of Automatic Control, Lund Institute of Technology, Lund, Sweden.

ÅSTRÖM K. J., J. J. ANTON, and K.-E. ÅRZÉN (1986): "Expert Control," *Automatica*, **22**, 3, 277–286.

ÅSTRÖM K. J. and T. HÄGGLUND (1984): "Automatic tuning of simple regulators with specifications on phase and amplitude margins," *Automatica*, **20**, 645–651.

ÅSTRÖM K. J. and T. HÄGGLUND (1988): *Automatic Tuning of PID Regulators*, ISA, Research Triangle Park, NC, USA.

Appendix A.

A.1 Properties of systems with $G(s) = k_p e^{-sL}/(1+sT)$

The ultimate frequency is defined by

$$\omega_u L + \arctan \omega_u T = \pi \qquad (A.1)$$

and the ultimate gain is given by

$$k_u k_p = \sqrt{1 + \omega_u^2 T^2}. \qquad (A.2)$$

Hence

$$\omega_u T = \sqrt{\kappa^2 - 1} \qquad (A.3)$$

and

$$\omega_u L = \pi - \arctan \omega_u T. \qquad (A.4)$$

Introducing θ we get

$$\theta = \frac{\omega_u L}{\omega_u T} = \frac{\pi - \arctan \sqrt{\kappa^2 - 1}}{\sqrt{\kappa^2 - 1}}, \qquad (A.5)$$

which is the exact relation between κ and θ.

A.2 Properties of systems with $G(s) = k_p 1/(1+s)^n$

The impulse response is

$$h_n(t) = k_p \frac{t^{n-1}}{(n-1)!} e^{-t} \tag{A.6}$$

which has maximum

$$\max h_n(t) = \frac{k_p (n-1)^{n-2}}{(n-2)!} e^{-n+1} \tag{A.7}$$

at $t_n = n - 1$. The step response $H_n(t)$ satisfies the relation

$$H_n(t) = H_{n-1}(t) - h_n(t). \tag{A.8}$$

Hence

$$H_n(t) = 1 - \sum_{i=1}^{n} h_i(t). \tag{A.9}$$

Furthermore

$$T_n = \frac{1}{h_n(n-1)} = \frac{(n-2)!}{k_p (n-1)^{n-2}} e^{n-1} \tag{A.10}$$

$$L_n = n - 1 - \frac{H_n(n-1)}{h_n(n-1)} = n - T + \frac{\sum_{i=1}^{n-1} h_i(n-1)}{h_n(n-1)}. \tag{A.11}$$

The ultimate frequency is given by

$$\begin{aligned} n \arctan \omega_u &= \pi \\ \omega_u &= \tan \frac{\pi}{n}, \end{aligned} \tag{A.12}$$

and the ultimate gain is

$$k_u = \frac{(1 + \omega_u^2)^{\frac{n}{2}}}{k_p}. \tag{A.13}$$

Numerical values for a few values of n are given in the following table.

n	L	T	θ	κ	$\kappa\theta$
1	0	1	0	∞	–
2	0.282	2.718	0.104	∞	∞
3	0.806	3.69	0.218	8	1.74
4	1.42	4.46	0.318	4	1.27
5	2.10	5.12	0.410	2.88	1.18
6	2.81	5.70	0.493	2.37	1.16
8	4.31	6.71	0.642	1.88	1.21

A.3 Properties of systems with $G(s) = k_p(1 - \alpha s)/(1 + s)^3$

The impulse response of the system is

$$h(t) = \frac{t^2}{2}e^{-t} - \alpha\frac{d}{dt}(\frac{t^2}{2}e^{-t}) = [(1-\alpha)\frac{t^2}{2} - \alpha t]e^{-t}. \qquad (A.13)$$

This has its maximum for

$$t_o = \frac{1 + 2\alpha + \sqrt{1 + 2\alpha + 2\alpha^2}}{1+\alpha}. \qquad (A.14)$$

The step response is

$$H(t) = 1 - e^{-t}(t + \frac{t^2}{2} + \frac{\alpha t^2}{2}). \qquad (A.15)$$

Hence

$$\begin{aligned} T &= \frac{1}{h(t_o)} \\ L &= t_o - \frac{H(t_o)}{h(t_o)} = \frac{t_o h(t_o) - H(t_o)}{h(t_o)} \end{aligned} \qquad (A.16)$$

and

$$\begin{aligned} \theta &= t_o h(t_o) - H(t_o) = \\ &= [\frac{t_o^3}{2}(1+\alpha) + \frac{t_o^2}{2}(1-\alpha) + t_o + 1]e^{-t_o} - 1. \end{aligned} \qquad (A.17)$$

The characteristic equation of the closed loop system is

$$s^3 + 3s^2 + (3 - \alpha k k_p)s + 1 + kk_p = 0. \qquad (A.18)$$

This equation has roots $\pm i\omega_n$ for the ultimate gain $\kappa = kk_p$. Hence

$$\begin{aligned} \omega_u^2 &= 3 - \alpha\kappa \\ 3\omega_u^2 &= 1 + \kappa. \end{aligned} \qquad (A.19)$$

Hence

$$\begin{aligned} \kappa &= \frac{8}{1+3\alpha} \\ \omega_u &= \sqrt{\frac{3+\alpha}{1+3\alpha}}. \end{aligned} \qquad (A.20)$$